Adaptive classification of EEG for dementia diagnosis

Ing. Matouš Cejnek

Supervisor: Doc. Ing. Ivo Bukovský Ph.D.

Abstract:

The paper presents new approach to dementia detection in time series of measured EEG. The proposed method introduced in this paper evaluates EEG signal according to included novelty. This novelty is estimated from prediction error and increment of adaptive weights obtained during prediction of measured EEG. Linear dynamic neuron was used as a predictor.

Keywords:

Novelty Detection, dementia, EEG, Gradient Descent, Linear Neural Unit

1. Introduction

Detection on dementia in its early stages is important for establishing of correct treatment. Early diagnosis and treatment could slow down the process of dementia development [1]. That is reason why pursuing of easily applicable methods for diagnosis is still actual topic. As a platform with good potential it seems to be EEG measuring and analysis. The EEG is relatively easy to measure and it were proven in history multiple times, that the signs of dementia are in EEG signal included. This uses the correlation between the degree of EEG abnormality and cognitive impairment [2][3]

Other methods of detection dementia from EEG are methods based on Probability Density Function of the Zero-crossing Intervals [4], Approximate Entropy [5], Fractal Dimension of the EEG [4].

The proposed method is based on Novelty Detection (ND) published recently in [6]. The ND method is unique, that contrary to other similar methods (Learning Entropy [7]), is using also error of the predictive model and not just increment of adaptive weights. Interesting advantage of the ND is, that even if the signal is complicated and dynamic, the used prediction model could be linear. That was found useful for ECG signal analysis [6]. Where it was proven, that even incorrectly chosen model could be sufficient for successful search for perturbations with ND.

The structure of this paper is following. In section 2 the used methods are introduced. After, in section 3 it is described the data and method validation. The results are summarized in section 4. And the discussion of the results follows in section 5.

2. Used Methods

The used methods are described in this section. For signal prediction we used a linear neuron as a model, adapted with Gradient Descent (GD) method [7]. The adaptation process of neural unit will be more described in following subsection 2.1. In subsection after 2.2, it is described the method of ND coefficient estimation, what was used for evaluation of every single sample of EEG data.

2.1 Adaptive Prediction

As was mentioned before, the method used for adaptation of predictive model is GD [7]. Input vector for prediction of every new sample is

$$\boldsymbol{x}(k) = [1\ y_r(k)\ y_r(k-1)\ ...\ y_r(k-4)]^T, \tag{1}$$

where vector *x(k)* stands for is the input vector and y_r is history of measured EEG values. Input vector contains history of *n* last samples and bias (in this case bias=*1*). For reason of getting better results, the measured time series is standardize according to equation

$$y_r \leftarrow Z_3(y_r) = \frac{y_r - \bar{y_r}}{3 \cdot \sigma_{yr}}. \tag{2}$$

With this standardization is possible to achieve better simulation stability, what leads to possibility of higher learning rate usage.

For even greater improvement of model stability a learning rate of GD (μ) adaptation is also used. Because when the learning rate is high in combination with high input values, the output of predictor will not converge. For such adaptation we use normalized learning rate [7]. How the method works is described in following equation

$$\eta = \frac{\mu}{1 + \boldsymbol{x}(k)^T \cdot \boldsymbol{x}(k)}, \tag{3}$$

where η is normalized replacement for the μ. Better stability of simulation allow us to use greater default learning rate, that in this case means better results of proposed method. Learning rate adaptation is evaluated before prediction of every single sample.

2.2 Novelty Detection Coefficient

For classification of measured EEG we use Novelty Detection based method [6]. The way how evaluate the ND is described in this subsection and the way how we use this method is the main achievement of this work. Coefficient of ND is estimated for every sample in measured data according to following equations

$$ND_{array}(k) = |e(k) \cdot \Delta w_i(k)| = |e^2 \cdot x \cdot \eta| = |\frac{e^2 \cdot x \cdot \mu}{1 + x(k)^T \cdot x(k)}|; i = 0...n, \tag{4}$$

$$ND(k) = max(ND_{array}(k)) \tag{5}$$

where *e* is prediction error, Δw_i is adaptation increment of of *i*-th weight of adaptive model. This is main principle of Novelty Detection (ND).

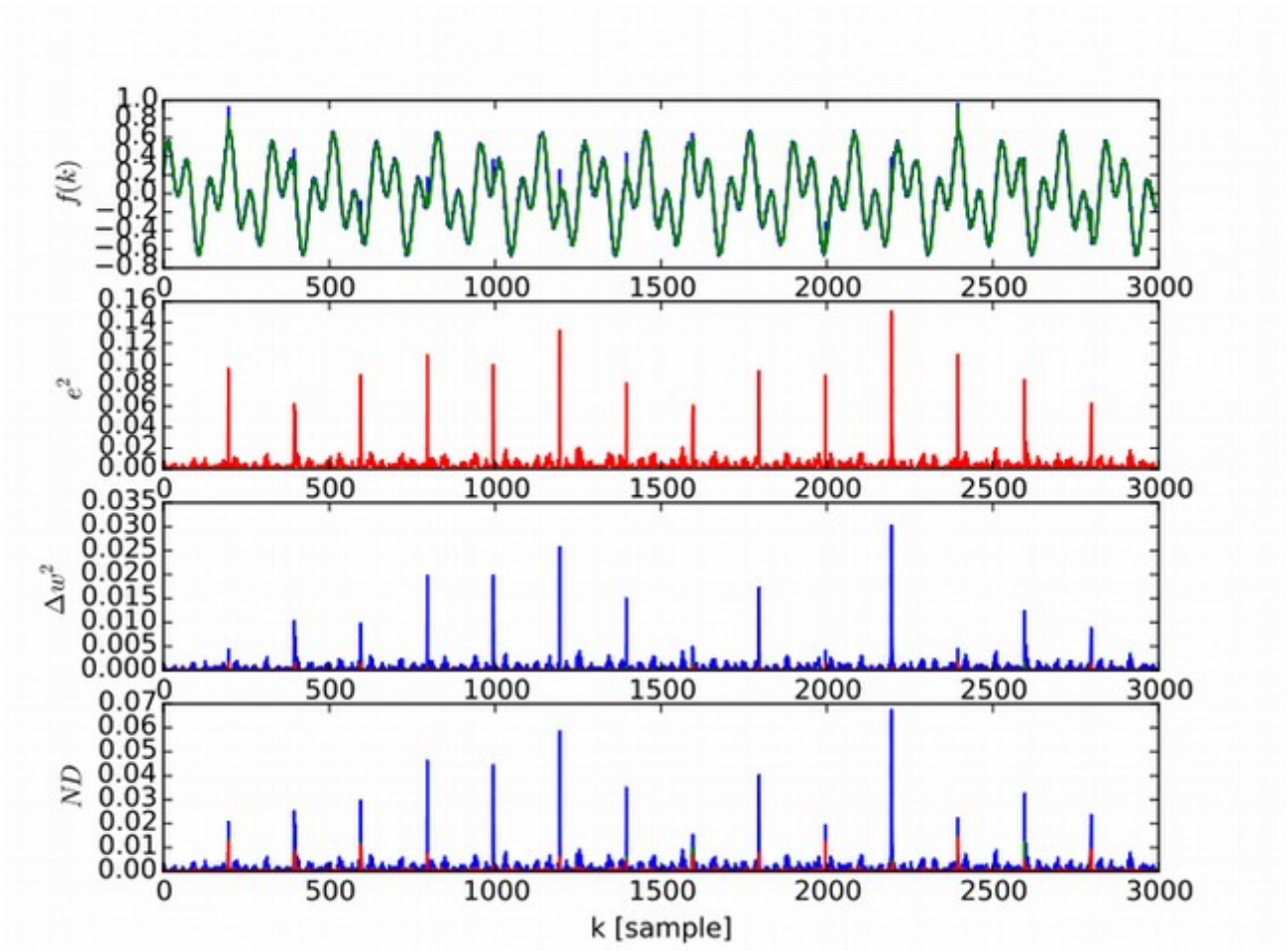

Fig. 1 *ND demonstrated on artificial data. From top: prediction result, absolute value of prediction error (red color), absolute value of adaptive weights increment, value of ND (equation (4))*

An example of ND estimation follows. As data for example was generated sum of two sine waves with periods of 21 and 53 samples. Into that time series a four small perturbations were introduced (located every 200 samples). Output of prediction and estimated ND is plotted in Fig. 1. As we can see, the introduced perturbations are in some cases more highlighted by square error and in other cases they are more obvious from square increment of weighs. The ND works as approximation of those attributes of predictive model. Therefore in this case it is better to describe the novelty in data with ND than other model features. For patient classification was used the standard deviation of the ND coefficients and entropy of the ND coefficients of patient EEG records.

3. Experimental Analysis

In this section will be introduced the implementation of ND in the way how we use it in this work. In subsection 3.1 used EEG data will be described. Results of prediction will be introduced in 3.2. And at the end of this section - subsection 3.4 will be described the validation of this proposed method.

3.1 Data Description

As a data we used records of EEG obtained from hospital. These records are manually selected section with no artefacts. Data selection contains records from 220 anonymous patients. From that selection a 110 patients match the clinical criteria of dementia and the rest are normal. Every patient has 2 to 5 manually selected records with length 90s or less.

3.2 Prediction

As we mention before, a linear neural unit with GD adaptation was used. EEG signal history for adaptation was really short (4 samples back and 9 samples back). This history cannot contain complete information about signal dynamics. That cause significant prediction inaccuracy. But such inaccuracy is not an issue for this method. Actually method works better with less precise simple model, than with more complicated models what we have tried. This fact could be an advantage in case of need of high speed implementation. Smaller amount of inputs means less calculations for one sample prediction. Prediction inaccuracy is shown in Fig. 2. The correlation coefficient between real measured EEG signal and prediction output is 0.3856 (4 samples history as input) and even just 0.3098 (9 samples history as input).

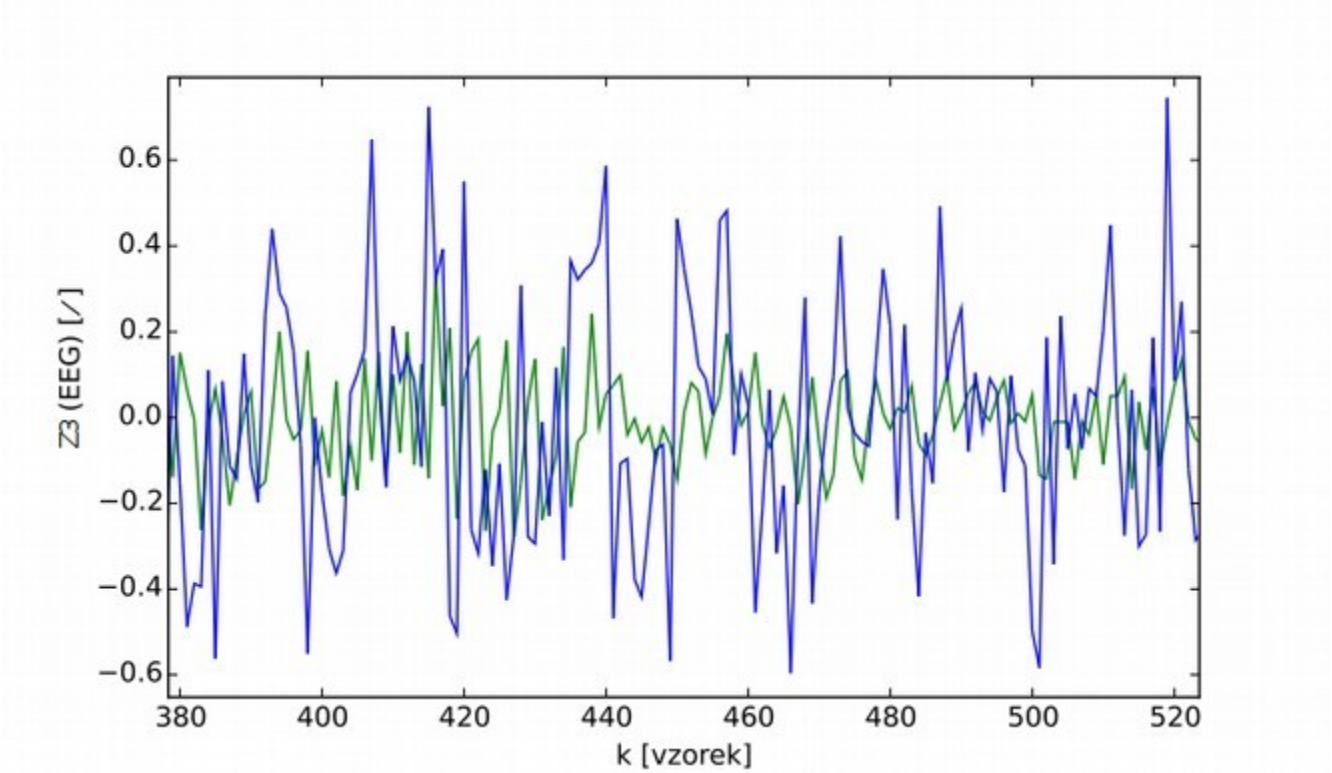

Fig. 2 *Prediction output of section of one EEG electrode (measured on patient with dementia), blue = standardize EEG, green = output of predictor. Predictor used history of 4 samples back.*

3.3 Evaluation of ND

The ND coefficient were obtained from attributes of predictive model (prediction error, increment of adaptive weights) for every EEG electrode for every patient. Two statistical functions applied on estimated ND were used to created criteria to decide, whether the EEG records belong to healthy person of patient with dementia. First investigated function was standard deviation and the second one was entropy. For every function a different length of history as an input of predictor was used – 4 samples back for standard deviation, 9 samples back for entropy. These settings gave as the best result.

Every patient has a multiple EEG records. From every record, we used data recorded by electrodes 13 to 19. Records has different lengths, but it does not matter, because we split the data into 1000 samples segments (7.8125 seconds). The ND coefficients of segment were estimated in third epoch of LNU learning. From the ND of every segment we estimated standard deviation and entropy. So we obtain multiple values for one patient (depends on lengths of patient records). For classification of the patient we use average of those values.

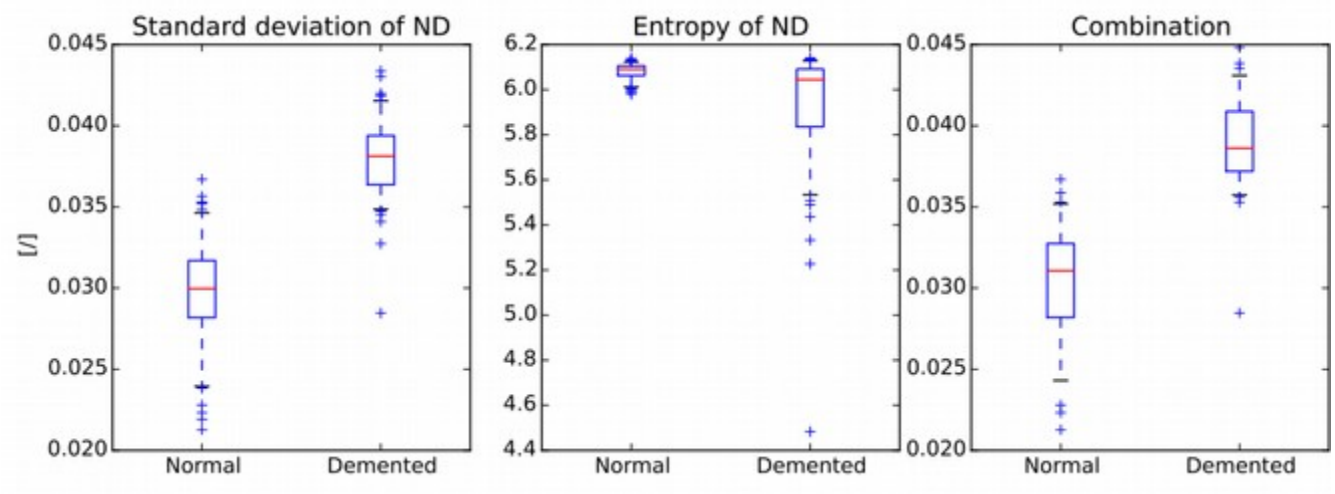

Fig. 3: *Box and whisker plots of results for all tested classification criteria*

3.4 Validation

For validation of method, we split patients between two groups. One group was for training (setting the criteria) and the other one for testing. Every group contains the same amount of patients with and without dementia. To eliminate the error caused by splitting into the groups, we split patients in groups randomly 100 times for every tested criteria. Average of all results was estimated and presented as a final result. We used three criteria for patient classification. The first one uses just standard deviation of ND, second works just with entropy of ND, and the last one uses both functions.

The criteria based on standard deviation was just median of all patients from training group. Patients from testing group with lower value than median was classified as healthy ones and those with higher value as patients with dementia. This criteria gave as specificity and sensitivity both over 93%. Distribution of values of this criteria is in Fig. 3.

The entropy based criteria was build on the finding, that demented patients have much bigger dispersion of entropy in ND than healthy patients (that is obvious from Fig. 3). So from training group was estimated lowest and highest value for normal patients and lowest and highest value of demented patients. Patients below lowest value of normal or above highest value of normal, were considered as demented. That means, that all demented patients with entropy in range of entropy dispersion of normal patients were marked incorrectly. That is reason, why this criteria has much lower sensitivity and specificity than first (standard deviation based) criteria. The sensitivity of this criteria is 82% and the specificity is 66%.

The last criteria uses both statistical functions. The main part of this criteria is standard deviation. But to that value, the penalization for entropy is added. If the patient has entropy in normal range of testing group, the penalization is 0. If the entropy is below the normal range, there is linear penalization, according to formula

$$P_i = \frac{C}{en_{normL} - en_{dementL}} \cdot (en_{normL} - en_i) \tag{6}$$

where P_i stands for penalization of i-th patient, en_{normL} is lowest value of normal patients from training group, $en_{dementL}$ is lowest value of demented patients from training group and en_i is value for i-th patient. This criteria has best classification results. The sensitivity and specificity are both 95%. Dispersion is shown in Fig. 3.

4. Conclusion

The novelty of EEG signal of 110 normal and 110 demented patients was estimated. Three different criteria for classification of patients were used. The best result was obtained with criteria what uses features of both other criteria. With method proposed in this paper and model settings what we used the specificity and sensitivity of 95% was achieved. Results of all criteria are summarized in following table

Tab. 1. *- Table of specificity and sensitivity for all tested criteria.*

Criteria based on	Specificity	Sensitivity
Standard deviation	93%	93%
Entropy	66%	82%
Standard deviation and entropy	95%	95%

And in the next table are the results of methods of other found and studied papers.

Tab. 2. *- Table of specificity and sensitivity of other methods.*

Method	Specificity	Sensitivity
Fractal Dimension Measure	99,9%	67%
Probability Density Function of the Zero-crossing Intervals	99,9%	78%
Approximate Entropy at P3	100%	70%
Approximate Entropy at P4	75%	80%
Other studies of American Academy of Neurology	70%	81%

5. Discussion

The validation results are good in comparison with results of other studied papers. But the questions is, how good the results will be for different EEG data (obtained from different machine, with different sampling and level of noise). It is also necessary to mention, that the used data were manually selected. This fact does not decrease the usability of the method, but method robustness should be also estimated in future for better knowledge of method applicability.

List of symbols

P_i	penalization of i-th patient	(1)
en	entropy of Novelty Detection coefficients	(1)
C	constant for estimation of patient penalization	(1)
ND	Novelty Detection	(1)
w	adaptive weights	(1)
Δw	increment of adaptive weights	(1)

x	input of neural unit	(1)
y_r	normalized input data	(1)
η	normalized learning rate	(1)
μ	default learning rate of gradient descent	(1)

Acknowledgment

The author and his supervisor would like to thank Prof. Aleš Procházka and his Digital Signal and Image Processing Research Group from the Institute of Chemical Technology in Prague for involving us in research of EEG data evaluation, for valuable consultancy, and for organizing Seminars Series on Digital Signal and Image Processing in Biomedical and Engineering Areas since 2014 that have accelerated and enriched our research effort.

Also, we would like to thank to Mudr. Oldřich Vyšata for consultancy and to the Department of Neurology, Faculty of Medicine in Hradec Králové, for providing us with anonymous EEG data sets.

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS15/189/OHK2/3T/12.

Literature

[1] G. Henderson, E. Ifeachor, N. Hudson, C. Goh, N. Outram, S. Wimalaratna, C. Del Percio, and F. Vecchio, “Development and assessment of methods for detecting dementia using the human electroencephalogram,” *IEEE Trans. Biomed. Eng.*, vol. 53, no. 8, pp. 1557–1568, Aug. 2006.

[2] R. P. Brenner, C. F. Reynolds, and R. F. Ulrich, “Diagnostic efficacy of computerized spectral versus visual EEG analysis in elderly normal, demented and depressed subjects,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 69, no. 2, pp. 110–117, Feb. 1988.

[3] D. W. Liddell, “Investigations of E.E.G. findings in presenile dementia,” *J. Neurol. Neurosurg. Psychiatry*, vol. 21, no. 3, pp. 173–176, Aug. 1958.

[4] G. Henderson, E. Ifeachor, N. Hudson, C. Goh, N. Outram, S. Wimalaratna, C. Del Percio, and F. Vecchio, “Development and assessment of methods for detecting dementia using the human electroencephalogram,” *IEEE Trans. Biomed. Eng.*, vol. 53, no. 8, pp. 1557–1568, Aug. 2006.

[5] D. Abásolo, R. Hornero, P. Espino, J. Poza, C. I. Sánchez, and R. de la Rosa, “Analysis of regularity in the EEG background activity of Alzheimer’s disease patients with Approximate Entropy,” *Clin. Neurophysiol.*, vol. 116, no. 8, pp. 1826–1834, Aug. 2005.

[6] M. Cejnek, P. M. Benes, and I. Bukovsky, “Another Adaptive Approach to Novelty Detection in Time Series,” 2014, pp. 341–351.

[7] I. Bukovsky, C. Oswald, M. Cejnek, and P. M. Benes, “Learning entropy for novelty detection a cognitive approach for adaptive filters,” in *Sensor Signal Processing for Defence (SSPD), 2014*, 2014, pp. 1–5.